图书情报工作
第65卷 第10期 2021年5月

融合用户兴趣及评论效用的评论信息推荐*

■ 聂卉　邱以菲
中山大学信息管理学院　广州 510275

摘　要：[目的/意义]Web 2.0 时代，在线评论质量参差不齐和过载问题十分严重，人们从中获取有价值内容的认知成本越来越高。探究以信息推荐方式解决评论过载的有效方案，以提升网络信息利用率和信息服务质量。论文提出的评论排名推荐方案关注评论的信息质量，更强调对用户的个人信息需求的满足。[方法/过程]研究运用概率主题模型，引入词向量构建主题空间下的用户模型和评论模型，通过将其纳入评论感知效用评测系统，实现融合用户兴趣和评论质量的评论推荐，推荐效果通过系统实验予以检测。[结果/结论]实验结果表明，评论信息质量和用户个体的信息需求，共同作用于用户对评论感知效用的满意度；推荐策略实现了二者的有机融合，三组不同推荐模式下的评测效果显示，相较于单纯的"兴趣推荐"和"效用推荐"，"融合推荐"综合满意度得分最高。

关键词：信息推荐　评论效用预测　用户建模　在线评论
分类号：P393.1　G202
DOI：10.13266/j.issn.0252-3116.2021.10.008

1 引言

移动互联时代，网上分享观点、发表评论成为人们的生活日常。在线评论提供的有价值内容，能帮助人们消除决策过程中的不确定性，对个体行为产生了深远影响。但在"人人都是创作者"的网络环境中，在线评论泛滥，质量参差不齐等问题变得日益严重，人们获取有价值内容的认知成本越来越高。为减轻用户认知负担，提升信息服务品质，各网站均采取了信息过滤机制。"淘宝"以是否有图、是否追评以及商品评级高低等作为筛选标准；"大众点评网"根据用户反馈屏蔽不可信内容，"豆瓣"和"亚马逊"等采用用户投票对评论排序。这些过滤策略主要针对信息质量，通过将高质量评论置顶来帮助用户快速获取有用信息。尽管如此，这些过滤策略没有关注到对用户个体需求的满足。

个体对信息的采纳，除受信息质量的影响，与个体信息需求有关，人们会更在意接收的信息中是否有其感兴趣的内容。特别是在信息量超出人认知负荷的情况下，人们快速浏览，希望尽快找到自己关注的内容。因此，以减轻用户认知负担，提升信息检索质量为目标来研究信息的效用价值，还应将用户个体要素纳入考虑中。个性化的信息推荐，是解决信息过载问题的重要途径。本文关注这一研究议题，探究将用户的个人信息需求融入评论效用预测模型的有效方案，以构建融合个人兴趣及评论质量两方面要素的个性化评论推荐系统。研究从推荐的视角来探讨信息服务精准化的实践策略，以期为控制评论信息泛滥，提升信息服务质量与信息利用价值的实践性工作提供参考。

2 相关研究

2.1 基于评论效用的评论排名与推荐研究

评论排名的实质是对评论效用进行评价，基于效用评价生成 TopN 推荐列表。近年研究中，郭顺利等[1]采用模糊层次分析法和加权灰色关联分析法预测评论效用，据之对评论排名，并选择信息量大的评论作最后推荐。张艳丰等[2]运用 K-means 算法对评论效用进行等级划分，继而对评论排序做优化。王忠群等[3]基于评论中"特征-观点"对的数量计算评论可信度对其排序，再以问卷方式邀请用户对 TopN 评论作评价。吴璠等[4]认为撰写人的历史评论可反映其发表评论的质量，他们基于撰写人曾经发表的评论对其建模，并融入评论模型，研究发现融入撰写人信息的模型能更好地

* 本文系国家社会科学基金项目"面向用户感知效用的在线评论的质量与控制研究"（项目编号：15BTQ067）研究成果之一。
作者简介：聂卉（ORCID：0000-0001-8567-3084），副教授，博士，E-mail：issnh@mail.sysu.edu.cn；邱以菲（ORCID：0000-0001-6114-6888），硕士研究生。
收稿日期：2020-09-07　修回日期：2021-01-08　本文起止页码：68-78　本文责任编辑：徐健

预测评论的效用。可见,评论的排名与推荐主要根据评价指标的计算。这些研究中,评价指标关注了评论的信息量、内容,可信度、撰写人水平,阅读群体整体感知效用等一系列要素,这些要素对辨识高质量评论起着至关重要的作用。

但最新的一份研究报告却指出,上述评价指标反映的只是评论信息在数据可靠性方面的质量,没有强调评论信息对目标信息用户的适用程度[5]。研究者认为对在线评论感知效用的评价是基于用户视角的一种信息质量评估,以用户的主观认识为出发点探讨信息的效用,需个体根据其个人经验对信息功能绩效进行系统性评估。因此,网络环境下的用户评论既要成为符合标准的高质量信息,更应关注评论信息对用户需求和期望的满足程度及为用户带来的价值[5]。持相同观点的研究者不乏其人,李健等[6]探究产品推荐,他们认为评论的效价应将消费者的个体偏好纳入考虑,寻找与消费者个人偏好匹配的高质量评论。E. Ben-abdallah 等[7]分析不同网络平台上的云服务评论的质量,通过计算评论者个人信息与云服务平台信息需求者的背景信息的相似性,来实现评论推荐。这些研究均以个性化的视角来研究评论的感知价值。

2.2 基于评论的推荐系统研究

2.2.1 基于评论的推荐系统分类

推荐是解决信息过载的有效方法,通过探知用户信息需求,推荐系统能够实现面向个人兴趣的信息推送,缓解过载信息造成的困扰[8]。产品推荐系统的核心是构建有效的用户和产品模型。评论信息因富含用户对产品的评价,从中提炼用户偏好构建用户模型,将其引入推荐系统成为近年的研究热点。L. Chen 等[9]从用户建模的视角,把相关研究划分为词项推荐,评分推荐和特征推荐三类。

基于词项的推荐归属内容推荐,直接利用评论文本对用户和产品建模。如 S. G. Esparza 等[10]从用户发表的评论中提取词项,以 TF-IDF 为词项权重生成用户模型,产品模型则基于目标产品的评论集,最后依据二者内容相似度进行推荐。耿立校等[11]的文献推荐系统基于用户阅读过的文献对其建模,用词向量表征词项,将用户和推荐目标(文献)的相似度计算升至语义层面。

评分推荐采用的协同推荐机制需要生成"用户－评分"矩阵,但矩阵稀疏性问题一直是协同推荐系统性能提升的瓶颈。解决方法之一是利用评论的文本数据预测用户对产品的评分,进而完善"用户－评分"矩阵,提升系统性能。W. Zhang 等[12]针对用户发表的产品评论,采用情感分析预测用户对产品的评分,构建基于"预测评分"的用户模型,用于产品推荐。C. Musat 等[13]通过凝练评论中蕴含的产品主题信息对用户评分加权,进一步了提升了模型的质量。张宜浩等[14]提出融合用户评分,情感倾向和产品内容的混合推荐算法,通过对稀疏的"用户－评分"矩阵进行填补修正,对产品进行排名推荐。

特征推荐聚焦评论内容中提及的推荐对象(产品)的细节。用户偏好描述直指产品特征,是基于产品特征评价的用户建模。H. Liu 等[15]的研究中,用户兴趣涉及关注度和需求度两项指标,关注度与产品特征提及率有关,需求度与产品评价有关,H. Liu 等从评论文本中提取产品特征词和评价词,构建了特征级用户兴趣模型,获得了更精准的产品推荐效果。张严亮等[16]在完善用户建模的研究中,关注用户兴趣的变化,他们借助用户发表的评论挖掘用户对产品各特征的评价,且定期更新评论,预测用户兴趣及变化趋势。基于特征的推荐同样采用协同推荐,需要对评论文本做细粒度分析。

2.2.2 基于评论的推荐系统建模方法

产品评论已成为推荐系统建模的重要数据源,从评论内容中提取用户关注的产品特征,作情感分析,预测用户对产品的评分是相关研究均涉及的问题。用户模型和产品模型主要采用隐性特征向量,概率主题模型和深度学习方法被广泛用于建模。Y. Bao 等[17]利用非负矩阵分解获得评论文本的隐主题,以主题分布反映用户偏好和产品特性。S. Feng 等[18]提出结合概率主题和随机漫步的混合建模思路,概率模型负责挖掘用户潜在偏好和产品隐特征,随机漫步用于构建全局的潜在关联。概率主题模型能够提取文本的语义特征,并通过降维缓解稀疏高维矩阵产生的问题。不足之处是忽略了重要的上下文信息。

深度学习模型能有效保留词序信息,更好地提炼文本特征,提升建模质量。L. Zheng 等[19]提出的 DeepCoNN 模型采用两个并行的卷积神经网络模型 CNN 从评论中学习用户和产品的隐向量表示,以评分预测函数为损失函数进行迭代训练,预测用户对产品的评分;S. Seo 等[20]在 CNN 的基础上,进一步引入注意力机制 ATT 来建模评论中的不同部分与用户偏好和产品特征的关联度,优化模型。C. Chen 等[21]为了充分利用了评分矩阵中用户及产品蕴含的信息,建模时,将评论文本和评分矩阵同时引入。冯兴杰等[22]借鉴 C. Chen 的

思路,将静态词向量换成可动态微调的 Bert 预训练模型,结合双向 GRU 结构和 ATT 机制分别从用户评论和产品评论中提取用户与产品的深层次特征向量,进一步提升推荐算法预测的准确率。

综上,推荐系统近年研究中,在线评论作为挖掘用户兴趣偏好的重要信息源受到重视。利用用户评论生成用户模型;或通过预测用户对产品的评分来提升"用户 - 评分"矩阵质量的协同推荐策略被普遍采纳。这些通过评论文本学习获得的用户和产品模型均以隐向量形式表征,概率主题模型和深度学习算法被广泛用于提升建模质量。

3 研究设计

本文以满足用户个性化信息需求为出发点,提出融合用户兴趣与评论信息质量的排名推荐策略,探究以信息推荐方式解决评论信息过载的有效方法。研究认为评论的效用在于为用户决策提供参考,而真正意义上的决策帮助,应面向用户的个人需求与偏好,人们对评论的感知效用因个体不同产生差异,评论推荐应更关注用户的个性化信息需求。李健等[6]从理论层面对这一问题进行了深度探讨;祝琳琳等[5]通过问卷获得数据,基于统计分析方法进行了实证性研究。E. Ben-abdallah 等[7]以信息需求者与评论撰写人的相似性作为信息推荐依据,但没对目标评论进行直接分析。本文则运用评论挖掘手段,通过构建个性化的评论感知效用评测模型,来实现个性化的评论信息推荐。

研究选择概率主题模型凝聚评论文本中的产品特征,生成用户和推荐对象(评论)的特征描述框架,评论文本直接用隐主题向量表征;用户兴趣模型则先通过交互手段获取其感兴趣的特征词,再映射至隐主题空间下,通过计算同一主题空间下用户兴趣与评论内容的语义相关度评测评论效用,相关度反映了评论信息与用户兴趣的一致性。最终推荐策略进一步将基于投票机制获得的评论感知效用值融入推荐指标计算,实现兼具评论质量的个性化评论推荐。本文研究将手机评论作为推荐对象,实验重点是用户兴趣建模和评论推荐策略的实施与评测,具体涉及 3 个问题:①如何对推荐的评论及用户兴趣进行有效建模?②如何构建融入用户兴趣的评论感知效用预测模型?③如何检测推荐策略的效果,验证推荐方案的科学性与合理性。

针对上述 3 个问题,整个研究分为 5 个模块,研究架构如图 1 所示。①数据预处理。对评论原始语料进行数据清洗和预处理操作,构建评论语料库。②构建评论 r 的主题模型。探索构建评论模型的合理方案。引入特征词典、特征同义词典,采用主题聚类生成评论内容的描述框架,通过挖掘产品关注点构建主题空间下的评论模型 r. topic_profile。③构建用户 u 的兴趣模型。基于主题模型描述用户兴趣,引入词向量,实现从特征词序列的用户模型 u. feature_profile 到主题模型 u. topic_profile 的映射。④推荐预测模型构建。构建融入用户兴趣的评论感知效用预测模型,设计评论排序推荐策略。⑤推荐效果评测。开发在线评测平台,采用"用户选择 - 评论推送 - 用户反馈"的交互方式获取用户对各推荐方案的评分,验测个性化评论推荐策略的实际效果。

5 个模块中,②③分别对应评论建模与用户建模,是评论个性化推荐系统的核心部分;④⑤将用户要素引入评论感知效用预测模型,重点探究个性化评论排名推荐策略及实现,并予以验证。

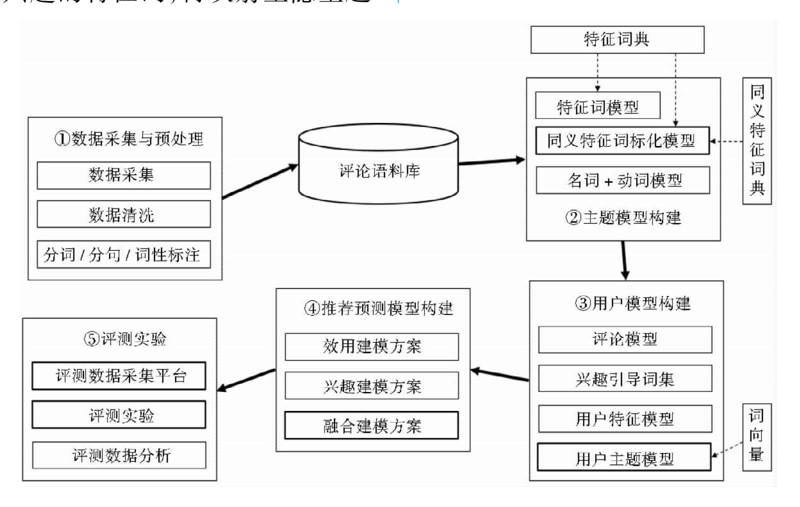


图 1 融合评论效用与用户兴趣的个性化评论排名推荐研究架构

4 模型与方法

4.1 主题建模

4.1.1 LDA 隐主题模型

研究的首要任务是在同一描述框架下对评论和用户建模,对此选择了 LDA 隐主题模型。LDA 是一个多层产生式概率模型,分词项、主题、文档 3 个层次。模型中,$D(|D|=N)$为文档集,$W(|W|=M)$为词集,$\varphi(|\varphi|=K)$表示隐主题。文档 $d \in D$ 由隐主题随机混合生成,一般表示成 W 上的序列模式 $d=\{w_i|w_i \in W\}$;隐主题 φ 则是 W 上的多项式分布。

α 和 β 是 LDA 模型的重要参数。α 反映文档集 D 中隐主题的相对强度,与主题的先验概率分布 Dirichlet 有关,β 刻画主题自身的概率分布。模型的主题层 φ 为"主题－词项"分布,由 $K \times M$ 的矩阵 β 参数化;文档层 θ 刻画"文档－主题"分布,由参数 α 确定;词项层 z_i($i=1,2,\cdots,M$)表示文档 d 分配在每个词项上的主题分量,服从 θ_d 的多项式分布。设定参数 α 和隐主题空间 φ,d 由两个过程生成:①从 Dirichlet($\theta|\alpha$)中随机选取一个 K 维向量 θ_d,产生文档 d 的主题分布。②依据条件概率 $p(w_i|\theta_d,\varphi_{1..K})$产生文档 d 的词项 。

LDA 模型的核心是推断"文档－主题"分布 α 和"主题－词项"分布 φ。在已知文档集词分布的前提下,逆向推导 $z_i(i=1,2,\cdots,M)$,进而获得 θ 和 φ。模型涉及多个未知量,一般采用近似解法,如常用的期望最大化算法(Expectation Maximization, EM)和 Gibbs 抽样。EM 算法[23]基于连续迭代使极大似然估计达到稳定,通过求解最大似然函数实现参数推断,本研究的 LDA 模型采用 EM 算法。

4.1.2 基于规范特征的主题模型

本研究选用 LDA 凝练评论内容,生成基于主题的特征描述框架。按照主题建模的常规做法,一般提取名词和动词等实词描述文档(后表述为"名词＋动词建模");但后续研究发现基于实词的聚类主题语义模糊,尤其是面向特定领域,于是引入领域特征词典,只取特征词进行聚类。这一优化策略可增强文档的主题表征力,缩减特征空间维度,提高计算效率,该方案被命名为 "特征词建模"。

但"特征词建模"仍是基于独立词项的建模方式,"屏幕""触摸屏""屏""大屏"等同义词均被视为不同词项,不能辨识词语间的同义关系。本研究在"特征词"基础上,对同义词进行规范,"触摸屏,屏,大屏"同义词等均用"屏幕"替换。这相当于对文档内容进行语义整合,如某评论的特征词描述为 d＝{屏,大小,彩屏,性价比,价格…}, 规范后变为 d＝{屏幕,尺寸,屏幕,价格,价格…}。经整合,文档表达升至概念层级。本文将其命名为"同义特征词标化"建模。

研究设计实验对比"名词＋动词","特征词"和"同义特征词标化"3 种建模方案的应用效果,选择最优方案作为评论和用户模型的描述框架。

4.2 主题空间下的评论及用户模型构建

4.2.1 评论模型

LDA 主题建模过程中,同步得到"文档－主题"概率分布矩阵 ,我们将 θ 表示成 $\text{Review_MAX}_{i\times k}$,I 对应评论语料集的文档数,K 为主题数。$\text{Review_MAX}_{i\times k}$的行向量即为评论 r 在主题空间下的概率分布描述,如公式 1:

$$\text{r.topic_profile}=[p_1,p_2,\cdots,p_k] \quad 公式(1)$$

4.2.2 用户兴趣模型

用户模型同样建立在隐主题空间上。为此,先设定一组描述用户兴趣的产品特征词 Interest_set,用户从中选择其关注的词项,算法将其选择的词项序列映射至隐主题空间。建模过程分 3 步:

步骤 1:设定 Interest_set,根据用户选择生成基于特征词的用户兴趣描述。

根据 LDA 聚类结果,同时参照电商平台对手机特征的分类,将描述手机性能的特征词划分为 8 个主题,分别为"屏幕效果、网络信号、外观设计、拍照摄影、影音娱乐、运行性能、性价比、电池续航",用户从中选择其关注的特征。如,用户 u 关注手机"外观"和"电池性能",则从对应主题中选取主题描述词表征 u,有 u. feature_profile＝{电池,续航,外观,外表,屏幕,机身,大小,…}。规范表达为公式(2)。其中, Topic(f)对应用户兴趣主题下的主题词集。将 u. feature_profile 映射至 LDA 隐主题空间。

$$\text{u.feature_profile}=\{t_i|t_i \in \text{Topic}(f), f \in \text{Interest_set}, i=1,2,\cdots,m\} \quad 公式(2)$$

步骤 2:用户兴趣的词向量表达。

词向量是基于浅层神经网络学习获得的词语的一种分布式表达,通过将词表示成一个 N 维高密度实数向量,词项对应了 N 维空间中的一点,点的间距反映了词项间潜在的语义关系。在将基于特征词的用户兴趣映射至主题空间之前,研究引入词向量,先将 u. feature_profile 转换为词向量矩阵 $\text{u.vec_MAX}_{m\times v}$($v$ 为词向量维度)。基于词向量描述的用户兴趣模型可传达语义,提高推荐准确率。而采用 $\text{u.vec_MAX}_{m\times v}$矩阵表示也便于

将用户模型映射至主题空间,用户兴趣和评论模型基于同一主题空间,即可看作空间中的两点,二者相关度直接用距离公式计算。研究引入的词向量为北京师范大学的开源中文预训练模型[24]。该词向量的训练语料为"百度百科",语料库规模 4.1G,向量空间维度为 300。

步骤 3:主题空间下的用户兴趣模型。

借由 LDA 聚类生成的"主题 - 词项"概率分布对主题 t 进行表达,如公式(3)所示:

$$t.feature_profile = \{ <f_i, w_i>, i = 1, 2, \cdots, n \} \quad 公式(3)$$

f_i 为描述主题 t 的特征词,w_i 为 f_i 的权重,n 为特征词个数。相应地,建立主题 t 的词向量矩阵 t. vec_MAX$_{nxv}$。在词向量空间下,将 u 的兴趣矩阵与主题 t 的转置矩阵相乘,同时纳入主题特征词权重矩阵 $W_{n \times v} = [w_1, w_2, \cdots, w_n]^T$,最终取矩阵运算结果的最大值为 u 与 t 的语义相关度(见公式 4)。根据公式 4 计算用户 u 与 K 个主题的相关度,生成主题空间下的用户兴趣模型,如公式 5 所示:

$$Sim_t = Max(u.vec_MAV_{m \times v} \times t.vec_MAX_{n \times v}^{T} \times W_{n \times l}) \quad 公式(4)$$

$$u.topic_profile = [Sim_1, Sim_2, \cdots, Sim_K] \quad 公式(5)$$

4.3 融合用户兴趣的评论排名与推荐

在用户模型和评论模型构建的基础上,将用户兴趣要素引入推荐模型。提出三种推荐模式,即基于用户兴趣的推荐(下称"兴趣推荐"),基于评论感知效用的推荐(下称"效用推荐")及融合用户兴趣及评论感知效用的推荐(下称"融合推荐")。

4.3.1 兴趣推荐

基于内容的文本信息推荐的本质是计算内容与用户兴趣的相关度,依相关度对推荐信息进行 TopN 排序。本研究用余弦距离测度二者相关度,令 $\vec{u}$ = u. topic_profile,$\vec{r}$ = r. topic_profile,计算见公式(6)。Pref_score 命名为"兴趣分"。

$$Pref_score = \vec{u} \cdot \vec{r} / (\| \vec{u} \| \times \| \vec{r} \|) \quad 公式(6)$$

4.3.2 效用推荐

效用推荐的排序指标直接采用基于网站用户投票的评论感知效用值,如公式(7)所示。Helpfulness_score 后简称"效用分"。

$$Helpfulness_score = \ln\left(\frac{Num\ of\ helpful\ votes}{Total\ votes}\right) \quad 公式(7)$$

4.3.3 融合推荐

融合方案将评论的兴趣分和效用分同时纳入计算公式 8,即融合用户兴趣和评论质量两方面要素的推荐策略,综合评价指标 Combined_score 命名为"结合分"。

$$Combined_score = w_p \cdot Pref_score + w_H \cdot Helpfulness_score \quad 公式(8)$$

式 8 中,w_p 和 w_H 分别为兴趣分和效用分的权重,满足 $w_p + w_H = 1$。具体应用中,两个权重设为参数,可对融合推荐方案进行灵活设置,该方案也将兴趣推荐和效用推荐纳入同一框架。

5 实验与分析

实验语料源自"中关村在线"(http://www.zol.com.cn)。构造 LDA 模型凝聚语料集中的产品的关注点,生成基于聚类主题的资源描述框架,即建立隐主题空间。用户兴趣通过开发数据采集平台获取,通过问卷反馈对融合推荐模式的应用效果予以检验。

5.1 数据与资源建设

5.1.1 评论数据及相关资源

语料采集时间为 2018 年 11 月。主要提取评论文本及评论的感知效用,共 14 506 条评论。相关资源还包括手机特征词表和同义词表,两个词表均为前期研究成果,特征词表汇总了手机特征词 301 个,同义词表包含 808 个词项。

5.1.2 用户兴趣数据和评测数据

依据研究设计开发用户数据采集和评测平台。评测页面列出预先设定的用户兴趣词集 ,受邀用户自主选择关注的产品特征,系统根据用户选择构建用户模型,生成评论推荐列表,向受邀用户进行评论推送;受邀用户需针对不同方案推送的评论进行评价,系统设置了信息适切度和信息质量两项指标。适切度测度用户个性需求方面的满意度,信息质量则是用户对评论的内容质量及形式的满意度评价。评价采用五级量表,评分越高,满意度越高。

课题组成员通过多种渠道发送评测系统链接,受邀者通过链接进入系统,按照指引完成兴趣特征词选择,并对推送评论作评价,系统后台汇总并分析评测数据。整个评测实验历时一个月(2019 年 11 月 - 2019 年 12 月),期间对平台各项功能进行了多次修正与完善,最终获得有效用户评测数据 208 份。评测平台的界面截图见图 2。

5.2 实验与结果

针对提出问题(参见 3 研究设计),共进行三组实验。实验 1:主题建模,从评论语料中提炼产品特征主



图 2　评论推送评测系统界面(评测界面截图)

题,对比“名词＋动词”“特征词”“同义特征词标化”三种建模方案的 LDA 聚类效果,确立模型描述框架;实验 2:用户建模,构建 LDA 隐主题空间下的用户模型;实验 3:评论推荐,对比“兴趣”“效用”“融合”3 种推荐策略的实际效果,检验个体要素对评论感知价值的影响。实验采用 python 语言实现,使用 jieba 进行中文预处理,利用 sklearn 的 Latent Dirichlet Allocation 类进行 LDA 建模,聚类可视化用 pyLDAvis 实现。

5.2.1　LDA 主题建模实验与结果

(1)模型参数和评测指标。为获取较优模型,实验首先考虑主题聚类的评测指标和参数设置。对于 LDA,主题间平均相似度最小时,聚类结构最稳定,模型最优。主题的平均相似度可用于测度模型稳定性。本研究用聚类主题间余弦距离的均值 Avg_similarity 测度聚类结构,该值越小,结构越稳定。另一指标为 *KL* 散度(Kullback-Leibler divergence),计算见公式(9)(10)。$KL(p \parallel q)$ 为主题 p 分布与主题 q 分布的 KL 散度,$p(x_i)$ 为 p 中词项的概率分布。因 *KL* 散度非对称,一般取 $KL(p \parallel q)$ 和 $KL(q \parallel p)$ 的均值作为 p 和 q 的 *KL* 散度。*KL* 散度测度分布差异,值越大,聚类区分度越大,聚类结构越好。

$$KL(p \parallel q) = \sum p(x_i) \log\left(\frac{p(x_i)}{q(x_i)}\right) \qquad \text{公式(9)}$$

$$KL(p,q) = \frac{KL(p \parallel q) + KL(q \parallel p)}{2} \qquad \text{公式(10)}$$

(2)模型参数设定。对于 LDA 模型,主题数 K 的取值十分关键,该参数与模型的 α 和 β 有关。研究将 K 作为优化参数,通过实验确定取值。图 3 为 3 种建模方案取不同 K 值的聚类效果。整体来看,随着 K 递增,Avg_similarity 呈下降趋势,说明主题间相似度减少,聚类结构稳定性增强。相反,*KL* 散度渐增,表明主题间差异拉大,内部凝聚力增强。而随着 K 递增,两个指标逐渐收敛。具体到三个建模方案,两组指标均显示,“同义特征词标化”的聚类效果明显优于“名词＋动词”和“特征词”。说明特征词典的引入有效减少了噪音词干扰,而特征词规范化进一步改善了矩阵的稀疏性,主题内容得以凝练,聚类模型质量最佳。因此,后续实验采用“同义特征词标化”的主题聚类方案。根据实验结果(见图 4)。两个评测指标均在 $K = 13$ 处开始趋缓,*KL* 散度 $= 8.267$,Avg_similarity $= 0.05$,最终确定 $K = 13$。

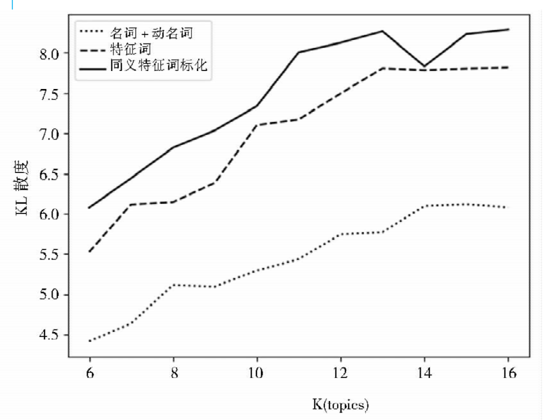


图 3　3 种特征建模方案下的 LDA 主题聚类效果
(左:纵坐标为 Avg_similarity;右:纵坐标为 KL 散度)

(3)聚类结果。图 5 为 $K = 13$ 时,用 pyLDAvis 生成的主题聚类效果图。整体观察,各主题分布均衡,大部分主题区分明确,少数出现交叠(主题 4 和 5,主题 1 和 2)。交叠易导致主题语义模糊,为此进行如下处理:对每个聚类主题,依词项分布概率从大到小对主题词排序,取前 8 个词项描述主题语义。若某词项同时出现在多个主题下,将其归属权值最高的聚类主题。如,“电池容量”均出现在主题 4 和主题 12 中,但在主

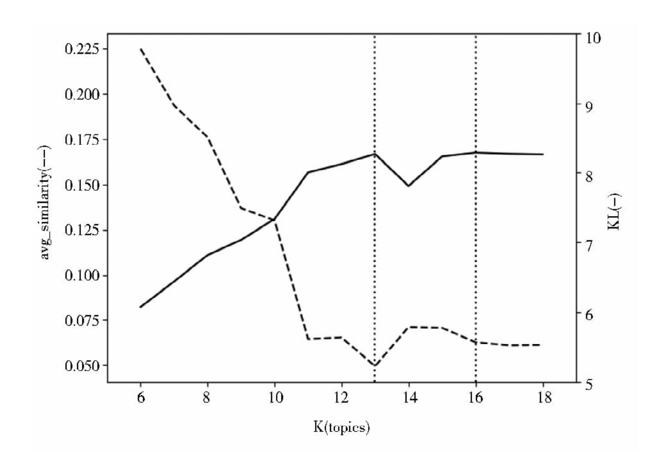


图 4 不同 *K* 值下的 LDA 主题聚类效果
(Avg_similarity 和 *KL* 散度)

题 12 下的权值(0.052)高于主题 4 下的权值(0.019),将其归在主题 12 下。聚类主题词经调整,可更好地明晰主题含义。根据各主题的主题词列表,参考数码网站对手机特征指标的设置,13 个主题分别对应到"运行性能、屏幕效果、网络信号、外观设计、拍照摄影、影音娱乐、性价比、电池续航、其他"9 个特征类下,据此生成用户兴趣选择特征词集 Interest_set,用于用户建模。

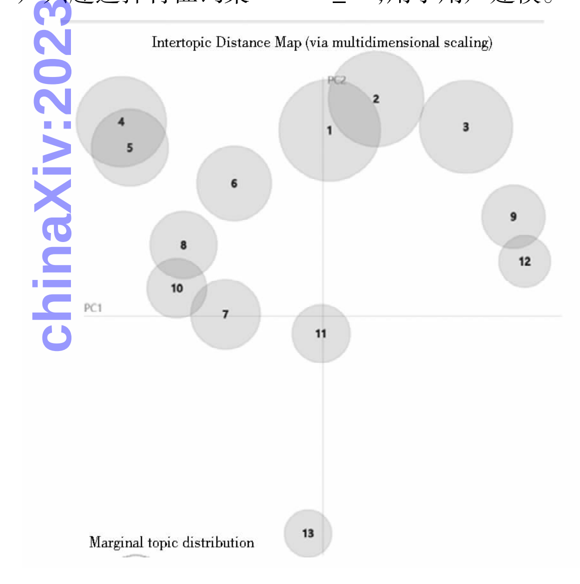


图 5 K=13 时的 LDA 主题可视化效果

5.2.2 用户兴趣建模实验

采用 4.2.2 的方法,用户从 Interest_set 中选择感兴趣的产品特征词,算法根据用户选择先生成 u.feature_profile,继而将 u.feature_profile 映射至主题空间下,生成 u.topic_profile。表 1 为用户建模示例,设用户从 Interest_set 选择了"屏幕效果"和"外观设计",根据主题词映射,生成 u.feature_profile = {外观,外表,屏幕,机身,大小,色彩,边框,…},继而引入词向量生成 u 在主题空间下的模型 u.topic_profile。表 1 所示,将特征词序列的用户模型映射至主题空间,相当于对描述用户兴趣的特征词进行语义扩展,用户信息需求表达更完善,能够匹配到更多与用户兴趣相关的评论,提升查全率。该模型的推荐效果在后续实验中得以检验。

表 1 用户 u 在隐主题空间下的兴趣表征(示例)

主题 topic	主题描述(排名前 8 的主题词,依权重排序)	u.topic_profile
1	手机 屏幕 外表 发热 使用 效果 游戏 东西	0.465
2	优点 时间 客服 待机 网络 上网 通话 网速	0.283
3	屏幕 外表 机身 边框 大小 色彩 金属 手感	0.733
4	像素 相机 屏幕 前置 拍照 后置 内存 电池容量	0.485
5	手机 拍照 效果 照片 外观 充电器 指纹 自拍	0.465
6	声音 通话 音乐 耳机 拍照 效果 外放 游戏	0.324
7	充电器 速度 运行 内存 指纹 发热 无线 软件	0.455
8	配件 品质 产品 处理器 品牌 国产机 性价比 样子	0.329
9	价钱 外观 性能 配件 屏幕 性价比 手感 定位	0.347
10	电池 续航 外观 能力 性价比 待机 使用 做工	0.434
11	系统 硬件 流畅 安卓 实用性 体验 下载 软件	0.279
12	流畅 运行 手感 外观 续航 系统 电池容量 像素	0.358
13	反应 指纹 关机 死机 图片 设置 重启 黑屏	0.191

5.2.3 融合用户兴趣的评论推荐实验

(1)评论排名序列差异分析。实验首先对 3 种推荐模式下评论排名进行差异性分析。设定用户选择的兴趣特征词(如"屏幕效果"),基于该兴趣词,分别采用 3 种推荐模式针对 4 款手机评论进行排名,就每款产品评论的 3 个排名序列进行 Friedman 检验,如表 2 所示。结果显示,对于 3 种推荐模式,各产品数据的 Friedman 统计量的概率 p 均小于 0.05,可认为 3 种模式下的评论排名存在显著差异,即相同兴趣词下,3 种推荐方案下的评论有不同的排序结果。根据平均秩,除产品 C 的评论,兼具用户个体要素和评论质量的"融合排名"模式与仅考量评论质量"效用排名"模式的差异明显,说明用户个体的兴趣偏好对评论感知价值具有一定影响力。

表 2 3 种推荐策略下评论推荐排名的 Friedman 检验
(用户兴趣特征词"屏幕效果")

产品	A (IPhoneX)	B (OPPOR9)	C (GalaxyS7)	D (Mate10)
评论量	167	1711	1035	74
基于兴趣排名的平均秩	1.921	1.833	1.660	1.923
基于效用排名的平均秩	1.893	1.947	2.175	1.802
基于融合排名的平均秩	2.192	2.231	2.172	2.286
Friedman 统计量	9.372	147.5	181.246	9.130
渐进显著性 p	.009 ***	.000 ***	.000 ***	.010 **

注:*表示 $p<0.1$;** 表示 $p<0.05$;*** 表示 $p<0.01$

实验进一步检验不同兴趣特征词下,评论推荐排名的差异。换言之,验证兴趣推荐策略能否满足个性化的信息需求,不同兴趣偏好的用户能否获得不同推荐。实验主要考量兴趣推荐和融合推荐两个方案,分别检测 8 个独立的兴趣特征词。检验结果显示(见表 3),对于不同的兴趣词,两个推荐模式的 Friedman 检验的显著性概率指标 p 均小于 0.05,可认为针对不同兴趣的用户,评论推荐排名存在显著差异。即不同兴趣偏好,会产生不同的推送结果。评论排序过程中,个体要素产生了一定作用,引入用户兴趣的推送随用户兴趣变化而不同。后续实验将对不同推荐模式的用户满意度做评测,评价结果可反映用户要素的影响力。

表 3　不同兴趣点下评论排名的 Friedman 检验(以 OPPOR9 为例,样本量 1711)

推荐模式	基于不同兴趣特征词的排名列表的秩均值								Friedman 统计量	渐近显著性 p
	屏幕效果	网络信号	外观设计	拍照摄影	影音娱乐	运行性能	性价比	电池续航		
兴趣推荐	4.600	4.091	4.201	3.933	4.224	4.927	5.125	4.936	405.610	.000 ***
融合推荐	4.723	4.123	4.128	4.394	3.847	5.098	4.725	5.002	410.37	.000 ***

注:*表示 $p<0.1$; ** 表示 $p<0.05$; *** 表示 $p < 0.01$

(2)3 种推荐策略的用户满意度分析。依照 5.1.2 构建评测平台(见图 2),发链接邀请用户参与实验。受邀者在产品匿名情况下选择感兴趣的产品特征词,平台依其选择推送一组评论,为确保数据质量,每一推荐方案依排名梯度推送三条评论(实验将排名 1、4、7 的评论进行推送)。同时,为屏蔽评论长度对评论感知效用的影响,推送评论长度控制在 30－460 个字符区间。评测者阅读评论后,对其适切度和质量评分并提交评测结果。实验用于检验用户个体要素对评论感知效用的影响,预期结果是:融入用户兴趣要素的推荐方案的满意度将高于单纯的"效用推荐"方案,"融合推荐"因同时考虑了信息质量,用户总体满意度最高。用户评分数据的统计结果如图 6 所示:

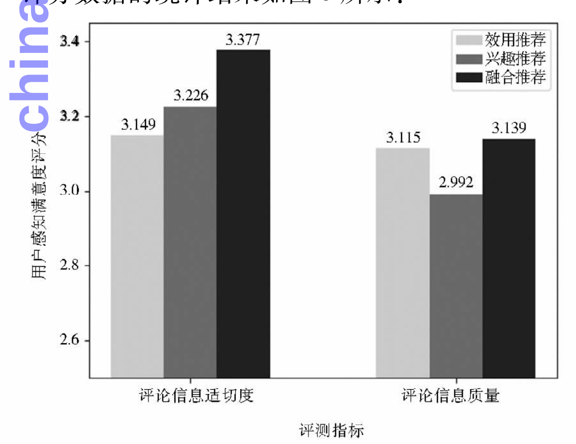


图 6　3 种推荐方案的用户满意度评价

如图 6 所示,对于信息适切度,即在满足用户信息需求方面,"融合推荐"获最高分;对于信息质量,"融合推荐"评分与"效用推荐"基本相当,该推荐模式依然可确保推送评论的质量。进一步作均值差异显著性检验,结果显示(见表 4),信息适切度方面,"融合推荐"与"效用推荐"的评价存在显著差异($t=-0.084$,$p<0.95$),与"兴趣推荐"的评价也存在显著差异($t=-2.228$,$p<0.05$),验证了"融合推荐"方案在满足用户信息需求方面的有效性。

评论质量的检验结果显示(见表5),"融合推荐"与"兴趣推荐"方案的评价存在显著差异($t=-2.2634$,$p<0.05$),"融合推荐"与"效用推荐"的差异不显著。进一步说明,"融合推荐"策略在信息质量方面表现较优,达到"效用推荐"方案同样的效果。但单纯基于"兴趣"的推荐若不考虑信息质量,推荐效果有限。

(3)推荐效果实例分析。综合上述实验结果,对评分最高的"融合推荐"方案进行重点分析。信息适切度方面,"融合推荐"较"效用推荐"的用户评分有显著提升(3.149→3.377)。原因在于"融合推荐"考虑了用户兴趣特征的语义匹配,这类评论中用户感兴趣的产品特征的评价内容更多,提升了信息的适切度。相较"兴趣推荐","融合推荐"则因有效过滤了兴趣特征提及率高但信息量过少的评论,一定程度上保障了信息质量,因而在三个推荐方案中表现最佳。推荐实例见表 6、表 7、表 8。

以表 6 为例,用户选择"屏幕效果"为兴趣词,"效用推荐"推送的评论中,没有"屏幕特征"的相关描述;"兴趣推荐"推送的评论中,出现了"屏幕""大小""6 寸"等相关特征,该评论有较高适切度,但因内容偏少,质量评价并不高,信息量是影响评论感知效用的重要因素;"融合推荐"推送的评论考虑了质量因素,强调了相关兴趣特征的提及率,表 6、表 7 实例均显示,该方案推送的评论在内容丰富度和满足个性化信息需求两方面都优于前二者。对于评论质量,"融合推荐"较"兴趣推荐"方案推送的评论质量评分有显著提升(2.992→3.139),原因在于"融合推荐"剔除了信息量少、效用低的评论,保障了推送信息的质量。

表 4 3 种推荐方案下的评论适切度用户评分差异 t 检验

配对样本		配对差值			t	自由度	Sig.
		平均值	标准偏差	标准误差平均值			
适切度	效用推荐 v 兴趣推荐	-0.077	0.979	0.068	-1.133	207	0.258
	效用推荐 v 融合推荐	-0.228	0.804	0.056	-4.084	207	0.000 ***
	兴趣推荐 v 融合推荐	-0.151	0.975	0.068	-2.228	207	0.027 **

注：** 表示 $p<0.05$；***表示 $p<0.01$

表 5 3 种推荐方案下的评论质量满意度评分差异 t 检验

配对样本		配对差值			t	自由度	Sig.
		平均值	标准偏差	标准误差平均值			
评论质量	效用推荐 v 兴趣推荐	0.123	0.992	0.069	1.793	207	0.074
	效用推荐 v 融合推荐	-0.024	0.762	0.053	-0.455	207	0.65
	兴趣推荐 v 融合推荐	-0.147	0.940	0.065	-2.263	207	0.025 * *

注：** 表示 $p<0.05$；***表示 $p<0.01$

表 6 产品 D 3 种排名方案的排名 No.1 评论对比分析(实例)

兴趣特征词	效用推荐	兴趣推荐	融合推荐
屏幕效果	配置杠杠的,就是 meta9 才出来已经那么贵了,meta10 的价格估计更离谱	mate10 好,功能强大。特别是屏幕大小合适,我用 mate7 是 6 寸的,用了多年,正合适	续航好音质不好 居然得抢 说了些有的没的就来说说手机吧,尺寸和我现在用的 P9 比起来稍微大上一小圈,但是屏幕毕竟大了近 1 寸,还是很震撼的。握在手里也觉得差不多,比 MATE9 要小一点。屏幕和 M9 都是 2K 的,显示很细腻…..

表 7 产品 B 3 种排名方案的排名 No.1 评论对比分析(实例)

兴趣特征词	效用推荐	兴趣推荐	融合推荐
运行性能	惊艳!!! 惊艳!!!! 太大了!!!! 太大了!!!! 只能当平板用!!!!	电池容量大,外观漂亮,续航出色价格高,性价比低,游戏运行不够流畅	屏幕大,外观漂亮,电池容量大,屏幕效果好,系统强悍,机身内存大,续航出色,分辨率高,处理器不错,运行流畅价格高,相机像素低,续航一般,音质一般,分辨率低,成像效果差,摄像效果一般,性价比低,游戏运行不够流畅。

表 8 产品 A 3 种排名方案的排名 No.3 评论对比分析(实例)

兴趣特征词	效用推荐	兴趣推荐	融合推荐
屏幕效果	期待是 iOS 11 + 苹果 A11 + 亮铜色 + 全面屏苹果在我心里是比较完美的,希望有双卡版苹果 iPhone X 难道不是之前曝光的苹果 8,会不会还有新款啊,不知道会不会还有新的配色,比如我喜欢的草绿色,粉色,真心希望价格不要太高,要不俺只能当吃瓜群众坐小板凳上看看发布会过过瘾了,如果性价比高的话一定入手一台,毕竟等了一年了。	上面的缺口明显影响屏幕的整体性,这样的全屏还不如不是;缺乏美感	便宜了将近 3000 块啊! 冲新全套配件银色 国行 256G 大内存在保修 359 天全套 7799 入手主屏尺寸 5.8 英寸 主屏材质 OLED, Multi-Touch 显示屏主屏分辨率 2436x1125 像素屏幕像素密度 463ppi 窄边框 4.57mm 屏幕占比 81.15% 实话实说,可以选择。

表 7、表 8 均显示,"融合推荐"推送的评论的内容较"兴趣推荐"方案推送的评论更丰富,能更好地帮助用户了解产品,消除决策中的不确定性,对用户感知信息价值产生了积极影响。综合其在适切度上的表现,"融合推荐"推送的评论既有较高的信息质量,又可满足用户的个性化信息需求。本文的研究假设得以验证。

6 研究结论与讨论

6.1 研究结论

本文探究以信息推荐方式解决评论信息过载的有效方案。研究采用概率主题模型,构建主题空间下的用户兴趣模型,并将其纳入评论感知价值计算模型,基于该模型提出融合用户兴趣和评论效用的评论推荐策略,通过在线评测系统,检验推荐策略的实效。整个研究围绕评论信息个性化推荐展开,研究贡献主要体现在 4 个方面：

(1)本研究构建了描述评论资源和用户兴趣的主题模型。建模算法引入领域特征词典和特征同义词典,采用标化后的评论特征词来描述评论文档,这一优化策略整合了同义词项,降低了文档特征维度,改善了矩阵的稀疏性,生成的主题模型解释度好,结构稳定,

被较好地用于评论和用户兴趣建模。

（2）本研究在语义层面构建用户兴趣模型。用户建模引入了词向量,对基于特征词序列用户模型进行语义表征,继而通过矩阵运算,将基于语义表征的用户模型映射至主题空间下。该建模方案将用户兴趣表达提升至语义空间。

（3）本研究将用户个性化要素引入评论评价机制,提出融合用户兴趣和评论效用的评论推荐策略。基于用户模型和评论模型,通过语义计算,面向用户预测评论兴趣得分,同时引入测度评论质量的评论感知效用。融合评论的兴趣得分和效用得分,提出两要素共同作用下的评论排名与推荐策略。

（4）本研究开发平台对推荐效果进行评测,对比分析“兴趣推荐”“效用推荐”和“融合推荐”3 个方案下用户对推送评论的评价。结果显示,“兴趣推荐”方案能较好地实现评论的个性化推荐,“融合推荐”方案在保障信息质量的同时,能够满足用户的个性化信息需求,综合表现最优。这说明,评论信息质量和用户的个性化信息需求,共同作用于用户对评论资讯的感知满意度,本文所提出的评论推荐策略,实现了二者的有机融合,从信息服务的视角来看,“融合推荐”无疑是更优的评论推荐模式。

6.2 实践启示

基于评论感知效用的评论排名推荐从用户群体感知出发,并没有关注到用户个体的信息需求。本研究验证了用户个体化要素对评论感知价值存在的显著影响力,提出了融合用户兴趣的评论推荐方案。这一研究能够帮助互联网企业更好地构建评论质量评测体系,为企业监控和有效利用网络信息提供理论借鉴和实践建议。例如,①鉴于信息质量和个性化因素对评论感知价值的显著作用,采用融合用户兴趣的信息推荐方式,构建基于用户需求并保证信息质量的信息过滤体系,优化信息服务模式。②面对目标用户,商家筛选兼具评论感知效用且满足其信息需求的评论予以呈现,能够提升用户的信息采纳率,促其产生购买行为,从而对商业运作产生积极影响。③对于评论信息服务及其他领域的资讯服务平台,引入检索机制获取用户信息需求,构建完整的资讯推荐系统,做到信息推荐精准化,可有效缓解信息过载带来的困扰,提升信息服务品质,促进信息服务社区的良性发展。

6.3 研究局限与进一步思考

本研究对建模和推荐方案设计得较简单。如,对于用户模型,表征用户兴趣的特征词采用等权重处理,但测试过程中发现,用户对产品性能的关注有侧重,后续研究可对描述用户兴趣的特征词设定权重,构建更精细的用户兴趣模型。同样,融合推荐方案中,评论质量因素和个性化要素的影响力也视为相等,但实验结果发现不同权重设置下的推荐结果存在较大差异,且不同类型的评论,两个因素的影响力也不尽相同,对这一问题,需采集更多类型的语料,作进一步探讨。后续研究也准备引入深度学习算法,对用户建模进行深入探索,从用户评论中提取用户特征,完善个性化推荐算法。

参考文献:

[1] 郭顺利,张向先,李中梅. 面向用户信息需求的移动 O2O 在线评论有用性排序模型研究——以美团为例[J]. 图书情报工作, 2015, 59(23) : 85－93.
[2] 张艳丰,李贺,彭丽徽,侯力铁. 基于情感语义特征抽取的在线评论有用性分类算法与应用[J]. 数据分析与知识发现, 2017, 1(12):74－83.
[3] 王忠群,吴东胜,蒋胜,等. 一种基于主流特征观点对的评论可信性排序研究[J]. 数据分析与知识发现, 2017, 1(10): 32－42.
[4] 吴璠,王中卿,周夏冰,等. 基于文本和用户信息的在线评论质量检测[J]. 中文信息学报, 2019, 33(9): 107－114, 140.
[5] 祝琳琳,李贺,刘金承,等. 在线评论信息质量感知评价指标体系构建研究[EB/OL]. [2020－12－12]. https://kns.cnki.net/KXReader/Detail? TIMESTAMP＝637449586862656250&DBCODE＝CJFQ&TABLEName＝CAPJLAST&FileName＝QBLL20201112002&RESULT＝1&SIGN＝MYZ8niwHlIVHm8g%2f8tmHWEKEjGs%3d.
[6] 李健,张军,苑清敏,等. 在线商品评论对消费者效用的改进分析——基于信息质量和消费者满意度理论视角[J]. 情报科学, 2018, 36(7): 137－144.
[7] BEN-ABDALLAH E, BOUKADI K, HAMMAMI M. Personalized cloud service review ranking approach based on probabilistic ontology[C] // Proceedings of 22nd international conference on business information systems. Seville: Springer, 2019:50－61.
[8] 姜霖, 张麒麟. 基于评论情感分析的个性化推荐策略研究——以豆瓣影评为例[J]. 情报理论与实践, 2017, 40(8): 99－104.
[9] CHEN L, CHEN G, WANG F. Recommender systems based on user reviews: the state of the art [J]. User modeling and user-adapted interaction, 2015, 25(2): 99－154.
[10] ESPARZA S G, P. O'MAHONY M, SMYTH B. Effective product recommendation using the real-time web[C] // Proceedings of the 30th SGAI international conference on innovative techniques and applications of artificial intelligence. Cambridge: Springer, 2010: 5－18.
[11] 耿立校,晋高杰,李亚函,等. 基于改进内容过滤算法的高校图

书馆文献资源个性化推荐研究[J]. 图书情报工作, 2018,62(21):112 – 117.

[12] ZHANG W, DING G, CHEN L, et al. Generating virtual ratings from Chinese reviews to augment online recommendations [J]. ACM transactions on intelligent systems and technology, 2013, 4(1):1 – 17.

[13] MUSAT C , LIANG Y, FALTINGS B. Recommendation using textual opinions[EB/OL].[2020 – 12 – 26]. https://www.researchgate.net/profile/Claudiu-Musat/publication/262348054_Recommendation_using_textual_opinions/links/57480c6308ae2301b0b8771f/Recommendation-using-textual-opinions.pdf.

[14] 张宜浩,朱小飞,徐传运,等. 基于用户评论的深度情感分析和多视图协同融合的混合推荐方法[J]. 计算机学报, 2019, 42(6): 1316 – 1333.

[15] LIU H, HE J, WANG T, et al. Combining user preferences and user opinions for accurate recommendation [J]. Electronic commerce research and applications, 2013,12(1): 14 – 23.

[16] 张炎亮,张超,李静. 基于动态用户画像标签的 KNN 分类推荐算法研究[J]. 情报科学, 2020, 38(8): 11 – 15.

[17] BAO Y, FANG H, ZHANG J. TopicMF: Simultaneously exploiting ratings and reviews for recommendation[C] //Proceedings of the 28th AAAI conference on artificial intelligence. Quebec: AI Access Foundation, 2014:2 – 8.

[18] FENG S, CAO J, WANG J. et al. Recommendations based on comprehensively exploiting the latent factors hidden in items' ratings and content [EB/OL]: [2020 – 12 – 26]. https://dl.acm.org/doi/pdf/10.1145/3003728.

[19] ZHENG L, NOROOZI V, YU P S. Joint deep modeling of users and items using reviews for recommendation[C]//Proceeding s of the 10th ACM international conference on web search and data mining. Cambridge, UK, 2017: 425 – 434.

[20] SEO S, HUANG J, YANG H, et al. Representation learning of users and items for review rating prediction using attention-based convolutional neural network [EB/OL]. [2020 – 12 – 26]. https://doogkong.github.io/2017/papers/paper8.pdf.

[21] CHEN C, ZHANG M, LIU Y, MA S. Neural attentional rating regression with review-level explanations [C] //Proceedings of the 2018 WWW conference on World Wide Web. Lyon, 2018: 1583 – 1592.

[22] 冯兴杰,曾云泽,崔桂颖. 基于近邻用户评论的推荐辅助网络[J]. 计算机应用研究, 2020, 37(10): 2956 – 2960.

[23] BLEI D M, NGA Y, JORDAN M I. Latent dirichlet allocation[J]. Journal of machine learning research. 2003, 3(4/5) :993 – 1022.

[24] 中文预训练模型[EB/OL]. [2020 – 12 – 20]. https://github.com/Embedding/Chinese-Word-Vectors.

作者贡献说明:

聂卉:研究问题提出,研究框架与研究流程设定,论文初稿,完善与修订;

邱以菲:数据采集,评测平台搭建与数据分析。

Integrating User Interests with Review Helpfulness for Review Recommendation

Nie Hui Qiu Yifei

School of Information Management, Sun Yat-sen University, Guangzhou 510275

Abstract: [**Purpose/significance**] In the Web2.0 era, the quality of online reviews is uneven and overloaded ; the cognitive cost of getting valuable content from them is increasing. In the paper, we proposed a review recommendation ranking scheme which focuses on the information quality of reviews and emphasizes more on the satisfaction of users' personal information needs. [**Method/process**] The probabilistic topic model was adopted in this study, user-profile and review model based on the topic model were generated by employing the Word2Vector. By incorporating them into review helpfulness evaluation system, the comment recommendation which integrates user interest and comment quality was realized. The models and related methods were tested and evaluated by a serial of systematic experiments. [**Result/conclusion**] The results indicate that both information quality and individual information interests are of influence on review perceived helpfulness. The "combining-specific" recommendation strategy, integrating the two factors effectively, performs better than the "interest-specific" and the "utility-specific" recommendation method. From the perspective of information service, the "combining" strategy should be with the high-priority for review recommendation.

Keywords: information recommendation review helpfulness prediction user-profile modeling on-line reviews